APPFL: OPEN-SOURCE FRAMEWORK FOR PRIVACY-PRESERVING FEDERATED LEARNING

KIBAEK KIM
Computational Mathematician
Mathematics and Computer Science
Argonne National Laboratory

MINSEOK RYU
YOUNGDAE KIM
Postdoctoral Appointees
Mathematics and Computer Science
Argonne National Laboratory

RAVI MADDURI
Computer Scientist
Data Science Learning
Argonne National Laboratory

June 2, 2022
APPFL: ARGONNE PRIVACY-PRESERVING FEDERATED LEARNING

Open-source PPFL framework

- **Motivations**
  - Distributed data from multiple institutions (e.g., Argonne, UChicago, etc.)
  - Avoid data transfer to a server
  - Potentially sensitive/private data

- **Science applications**
  - Biomedical data
  - Smart meters deployed in grid
  - Experimental facilities (e.g., APS X-ray beamlines)
  - National security (e.g., critical infrastructure)
FEDERATED LEARNING

- Machine learning without centralizing training data
  - No direct data sharing or storing

- More benefits
  - Learning a global/shared model
  - Utilizing a localized model at each client side
  - Personalization

- Two settings:
  - Cross-device FL (1000s and 1Ms of small devices)
  - Cross-silo FL (a few large data repositories)
PRIVACY-PRESERVING TECHNIQUES

- Some techniques in FL
  - Homomorphic encryption: limited to certain operations
  - Secure multi-party computation: computationally expensive
  - Differential privacy: potential accuracy loss

- Differential Privacy
  - The two outcomes are indistinguishable for all D1 and D2 which differ in one individual’s data.

\[
\ln \left( \frac{P(\mathcal{R}(D_1) \in S)}{P(\mathcal{R}(D_2) \in S)} \right) \leq \epsilon
\]
OVERVIEW: APPFL FRAMEWORK

Step 1 (→)
Server broadcasts global model parameters

Step 2
Each client computes local model parameters

Step 3 (→)
Each client sends local model parameters (randomized)

Step 4
Server updates global model parameters
MAJOR COMPONENTS OF APPFL

- **Training algorithms:**
  - IIADMM, FedAvg [McMahan et al., 2017], ICEADMM [Zhou and Li, 2021]
  - Any user-defined FL algorithms can be added.

- **Differential privacy:**
  - Random perturbation with Laplacian noises [Dwork et al., 2006]
  - More advanced schemes can be added.

- **Communication protocols:**
  - gRPC: communication between multiple platforms and languages
  - MPI: efficient communication in a cluster environment

- **User-defined model and data:**
  - Inherits PyTorch’s neural network module, torch.nn.Module
  - Dataset class that inherits the PyTorch’s Dataset
ADVANCED PPFL ALGORITHMS

Implementation of novel training algorithms

- (state-of-the-art) OutP: Inexact ADMM (IADMM) + output perturbation
- (APPFL) ObjP: IADMM + objective perturbation
- (APPFL) ObjPM: IADMM + objective perturbation + multiple local updates

![Graphs showing comparison between OutP, ObjP, and ObjPM with respect to testing error over iterations. Stronger privacy implies weaker learning, whereas stronger learning implies weaker privacy.](image)
USE CASE

Chest X-ray classification for COVID-19 cases

- FL can produce more accurate model, compared to the models trained on local datasets.
- DP is applied to protect chest X-ray data from reverse-engineering the model gradients communicated during training.
- Collaborations with UChicago Medical School and Broad Institute
USE CASE
Federated load forecasting for electric distribution system

Parameters with differential privacy

- Electricity consumption at household level
- Distributed data at a large number of devices
- Privacy concern
CONCLUDING REMARKS

- APPFL: open-source Python package with support of any Pytorch models
  - The package has been released (v0.2.0).
- Any user-defined ML model can be trained on decentralized data while ensuring data privacy.
- Customized PPFL algorithms can be easily implemented.
- Applications: national security, smart grid (Fig. 1), biomedicine, experiments (Fig. 2), etc.
- Collaborations and contributions are welcome!

Fig 1. Network architecture in smart grid (modified from https://doi.org/10.1016/j.comnet.2012.12.017)

Fig 2. Multiple experimental devices at APS
REFERENCES

- [https://github.com/APPFL](https://github.com/APPFL)
ACKNOWLEDGEMENT

Funding supports and Contributions

- DOE-ASCR PALISADE-X Project
- DOE Early Career Research Program (on data-driven optimization)

- Minseok Ryu (ANL)
- Youngdae Kim (former postdoc at ANL)
- Ravi Madduri (ANL)
- Jordan Fuhrman (UChicago)
- Nick Dodd (ASU)
THANK YOU